Nottingham Equation of State Toolkit: Statistical challenges and solutions

Richard Graham, Joanne Dunster, Martin Nelson, Simon Preston, Richard Wilkinson

School of Mathematical Sciences
University of Nottingham

Nottingham - May 2014

Talk Plan

- Recap
- Why optimisation fails
- Why vanilla MCMC fails and how to spot it
- Parallel tempering

Recap

Parametric function

$$P = f_\theta(V)$$

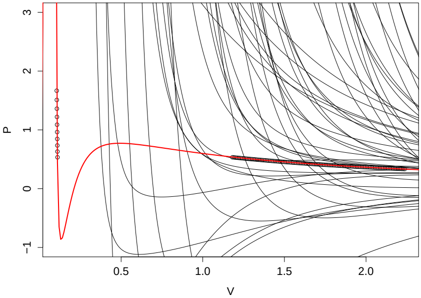
(or a $> 1d$ equivalent)

Data

$$\mathcal{D} = \{V_i, P_i\}_{i=1}^N$$

Prior region:

$$\theta \in \Theta$$



Recap

Parametric function

$$P = f_\theta(V)$$

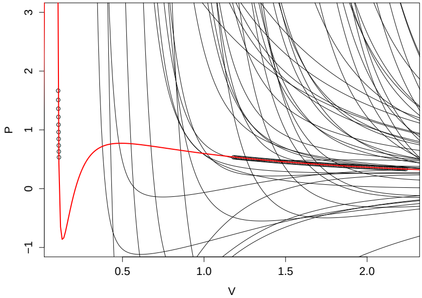
(or a $> 1d$ equivalent)

Data

$$\mathcal{D} = \{V_i, P_i\}_{i=1}^N$$

Prior region:

$$\theta \in \Theta$$



How do we find values of θ that lead to good matches between model predictions $f_\theta(V)$ and observations, P?

Estimating θ

One option is to maximise some objective function, e.g.,

$$S(\theta) = \sum_{i=1}^{N} (P_i - f_\theta(V_i))^2 + C(f_\theta)$$
$$\hat{\theta} = \arg\min_{\theta} S(\theta)$$

where $C(f_\theta)$ is any additional criteria on f, e.g., fugacity constraints.

Opimization is hard!

Consider the function

$$P = f_\theta(V) = \frac{T}{V + \theta_1} + \frac{\theta_2}{V^2 + \theta_3} + \frac{\theta_4}{V^3 + \theta_5} + \frac{\theta_6}{(V - \theta_7)^6}$$

Can we use a numerical optimiser to find good values of $\theta = (\theta_1, \ldots, \theta_7)$?

Opimization is hard!

Consider the function

$$P = f_\theta(V) = \frac{T}{V+\theta_1} + \frac{\theta_2}{V^2+\theta_3} + \frac{\theta_4}{V^3+\theta_5} + \frac{\theta_6}{(V-\theta_7)^6}$$

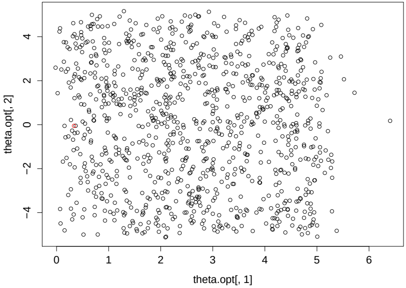
Can we use a numerical optimiser to find good values of $\theta = (\theta_1, \ldots, \theta_7)$?

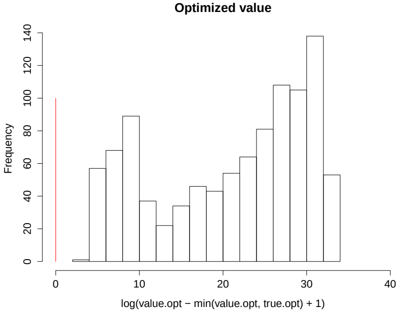
Experiment:

- Pick a 'true' value of θ and generate some P, V data.
- Pick 1000 random start points $\theta_i \in \Theta$
- Run an optimiser starting at each θ_i
- How often does the optimiser find the 'true' value?

Optimization is hard!

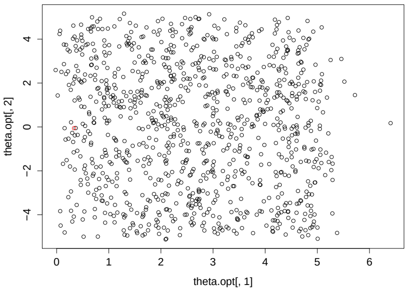
Standard optimisation fails badly and UQ fails as a result:

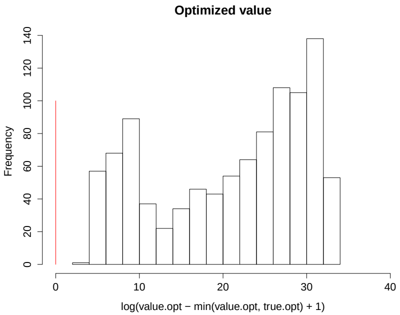


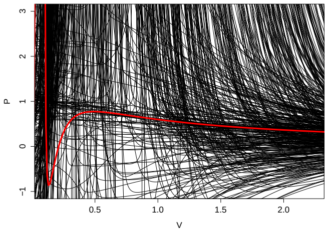


Optimization is hard!

Standard optimisation fails badly and UQ fails as a result:

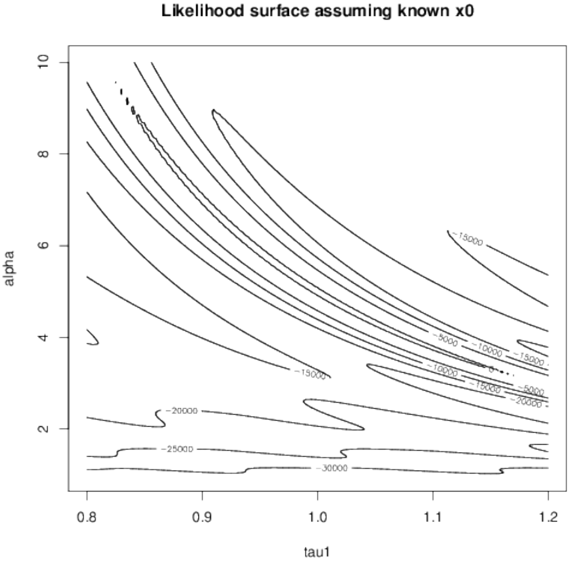






- The optimiser reported it had converged in every case.
- the classical/frequentist approach to statistics is thus almost impossible

Why does this fail?



Bayesian approach

Instead of optimizing, we seek to find the Bayesian posterior distribution

$$\pi(\theta|\mathcal{D}) \propto \pi(\theta)\pi(\mathcal{D}|\theta)$$

Bayesian approach

Instead of optimizing, we seek to find the Bayesian posterior distribution

$$\pi(\theta|\mathcal{D}) \propto \pi(\theta)\pi(\mathcal{D}|\theta)$$

Advantages:

- Meaningful uncertainty quantification
- Powerful algorithms exist for finding the posterior
- More realistic philosophical interpretation, e.g., no single 'correct' value assumed or required

Bayesian approach

Instead of optimizing, we seek to find the Bayesian posterior distribution

$$\pi(\theta|\mathcal{D}) \propto \pi(\theta)\pi(\mathcal{D}|\theta)$$

Advantages:

- Meaningful uncertainty quantification
- Powerful algorithms exist for finding the posterior
- More realistic philosophical interpretation, e.g., no single 'correct' value assumed or required

Disadvantages

- Inference algorithms are complex
- Additional modelling decisions needed

MCMC

We've seen already how we can use Markov chain Monte Carlo (MCMC) algorithms to find the posterior distribution $\pi(\theta|\mathcal{D})$

MCMC

We've seen already how we can use Markov chain Monte Carlo (MCMC) algorithms to find the posterior distribution $\pi(\theta|\mathcal{D})$

Simulate a Markov chain, $\theta_1, \theta_2, \theta_3, \ldots$ such that the stationary distribution is $\pi(\theta|\mathcal{D})$

- If currently at θ_n, propose a move to θ'

$$\theta' \sim q(\theta_n, \theta')$$

- Accept move with probability

$$r = \frac{\pi(\theta'|\mathcal{D})q(\theta', \theta_n)}{\pi(\theta_n|\mathcal{D})q(\theta_n, \theta')}$$

  and set $\theta_{n+1} = \theta'$

  Otherwise set $\theta_{n+1} = \theta_n$

MCMC

We've seen already how we can use Markov chain Monte Carlo (MCMC) algorithms to find the posterior distribution $\pi(\theta|\mathcal{D})$

Simulate a Markov chain, $\theta_1, \theta_2, \theta_3, \ldots$ such that the stationary distribution is $\pi(\theta|\mathcal{D})$

- If currently at θ_n, propose a move to θ'

$$\theta' \sim q(\theta_n, \theta')$$

- Accept move with probability

$$r = \frac{\pi(\theta'|\mathcal{D})q(\theta', \theta_n)}{\pi(\theta_n|\mathcal{D})q(\theta_n, \theta')}$$

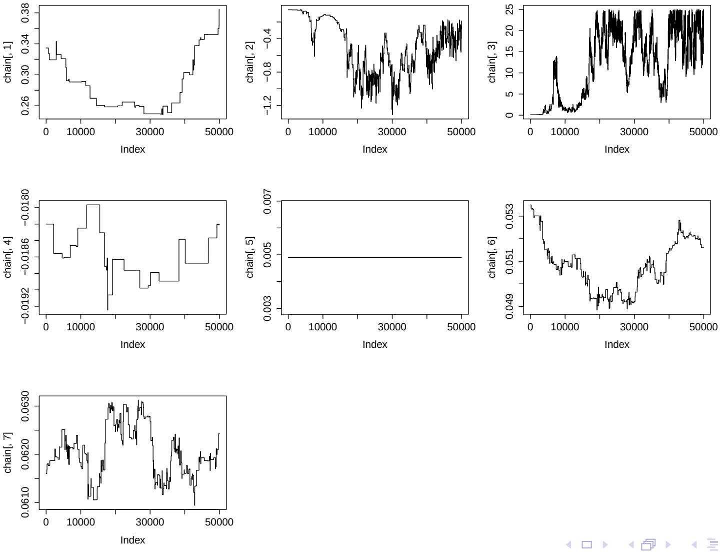
  and set $\theta_{n+1} = \theta'$

  Otherwise set $\theta_{n+1} = \theta_n$

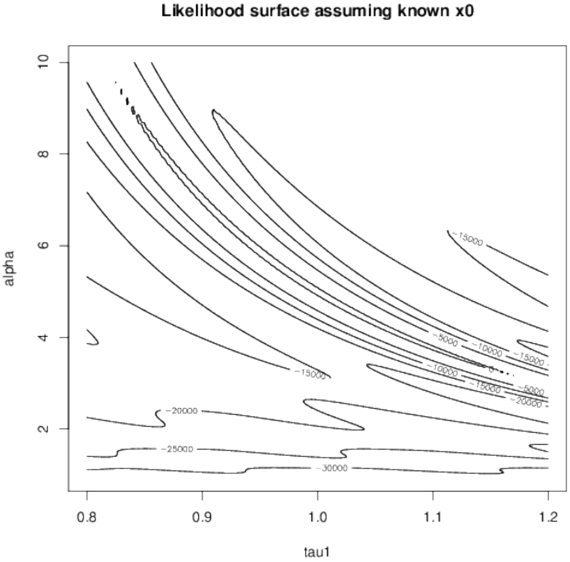
The art is in choosing a good q

No free lunch: poor $q \to$ poor results

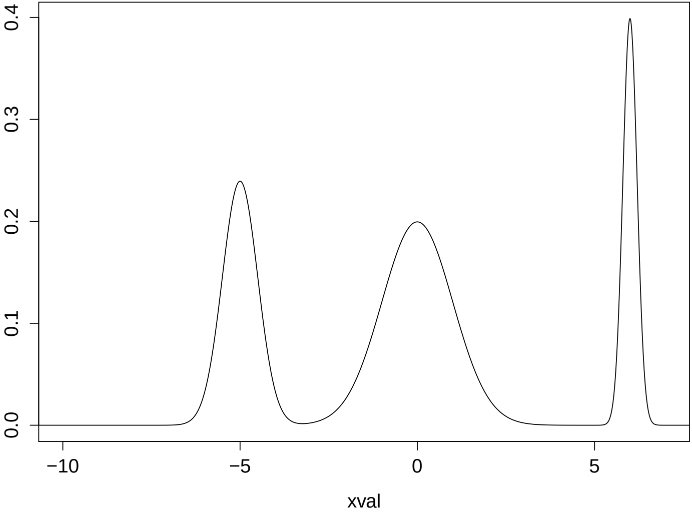
Trace plots can be used to diagnose problems with mixing and convergence



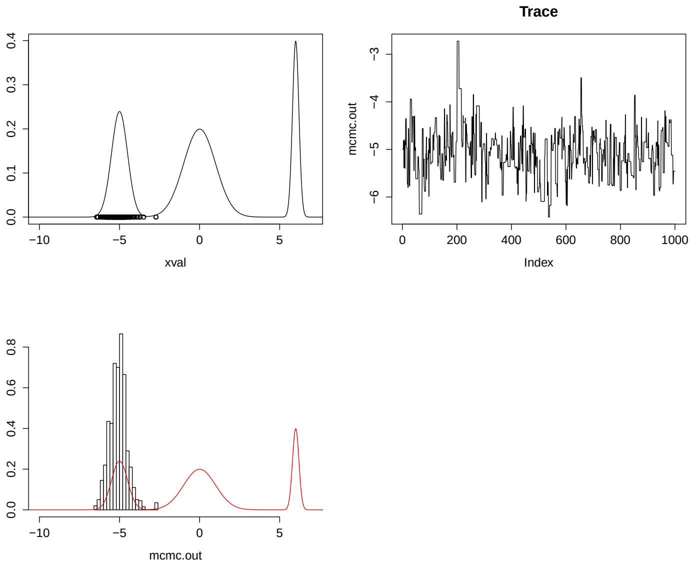
Why is this hard?



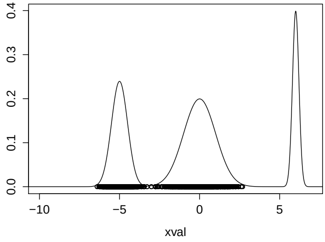
Simple 1d demo

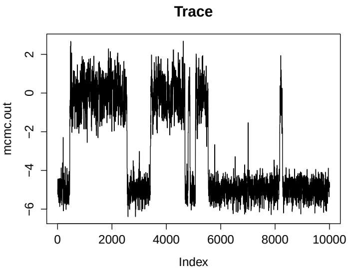


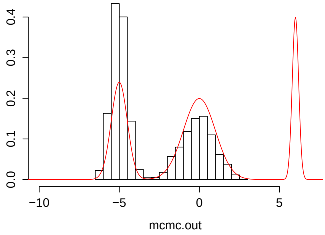
Simple 1d demo



Why is this hard?

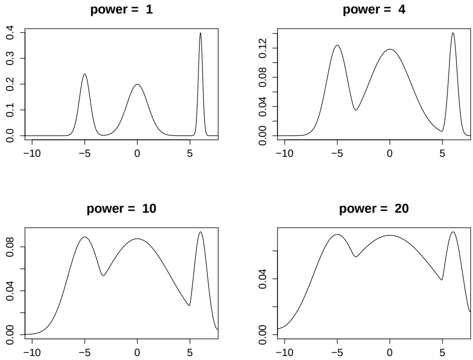






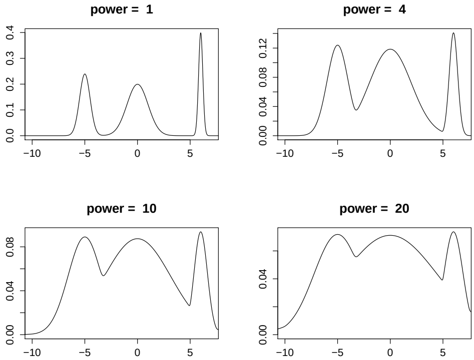
Parallel tempering

- Heating the likelihood makes it flatter and easier to explore
- Exploring $\pi(\mathcal{D}|\theta)^{\frac{1}{p}}$ with MCMC is easy if p is large. $p = 1$ corresponds to the desired posterior
- Larger powers p are thought of as 'hotter' temperatures



Parallel tempering

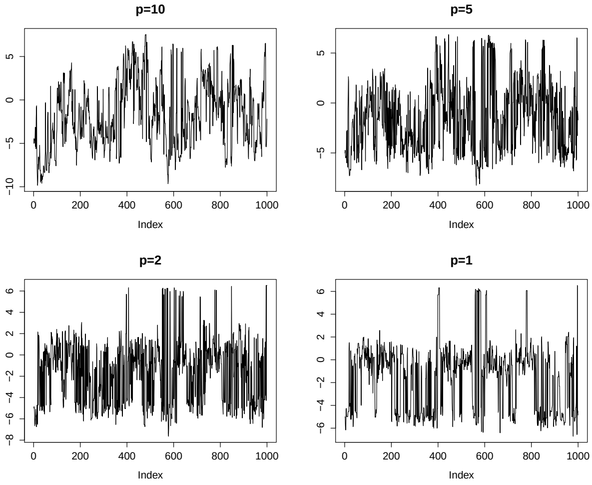
- Heating the likelihood makes it flatter and easier to explore
- Exploring $\pi(\mathcal{D}|\theta)^{\frac{1}{p}}$ with MCMC is easy if p is large. $p = 1$ corresponds to the desired posterior
- Larger powers p are thought of as 'hotter' temperatures



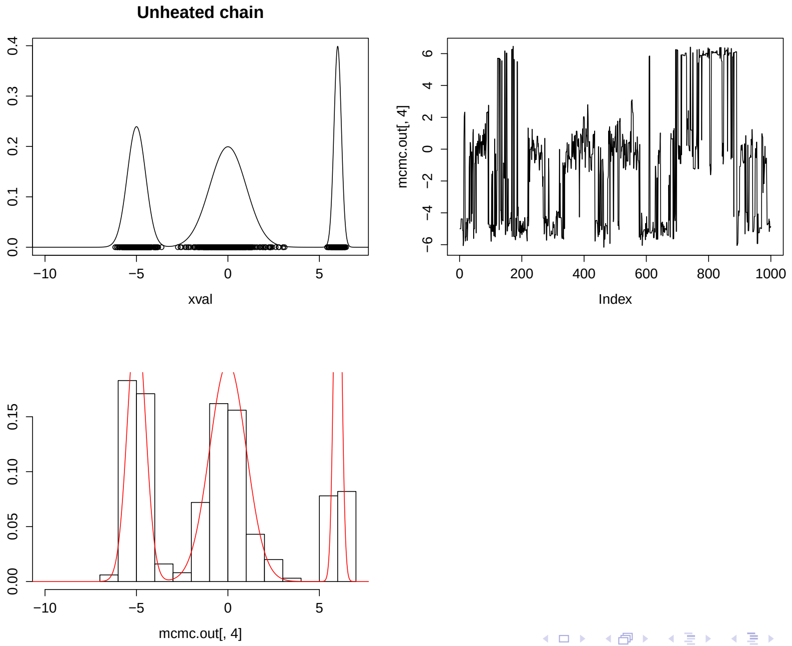
Idea:

- run multiple MCMC chains - some hot and some cold
- propose switches between the chains (maintaining detailed balance)

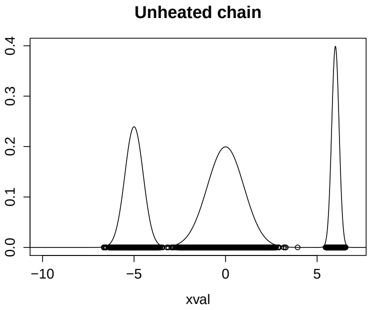
Parallel tempering

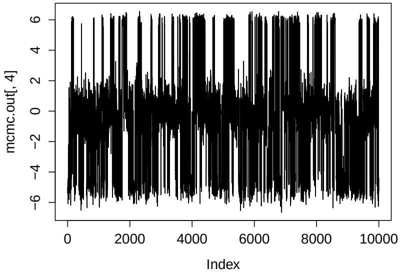


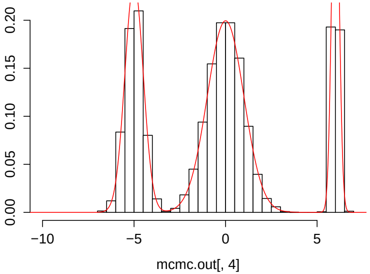
Parallel tempering - untuned



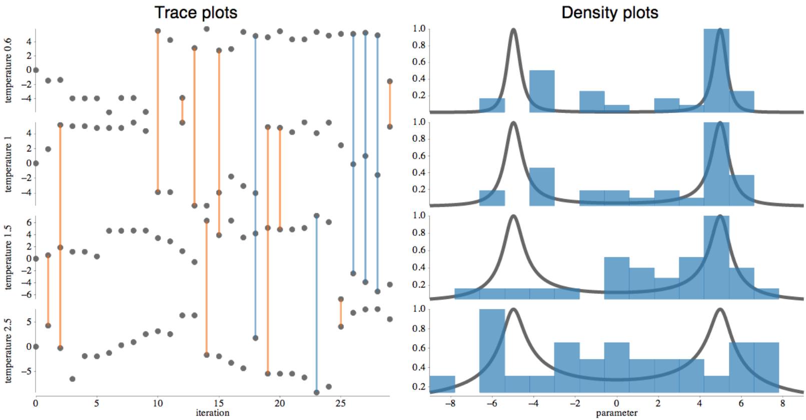
A longer run


Unheated chain
0.0
0.1
0.2
0.3
0.4
−10
−5
0
5
xval


mcmc.out[, 4]
−6
−4
−2
0
2
4
6
0
2000
4000
6000
8000
10000
Index


0.00
0.05
0.10
0.15
0.20
−10
−5
0
5
mcmc.out[, 4]

Parallel tempering



https://www.youtube.com/watch?v=J6FrNf5__G0

Boutique proposals, q

The key determinant of whether an MCMC sampler works well or not is the choice of proposal q.

Boutique proposals, q

The key determinant of whether an MCMC sampler works well or not is the choice of proposal q.

In the GUI we have a carefully tuned proposal that combines multiple different moves.

- With probability p_1 we update a single parameter using a Gaussian random walk proposal
- With probability p_2 we update a block of two parameters using a Gaussian random walk
- With probability p_3 we update all 7 parameters
- With probability p_4 we propose a swap between parameter values in chains of different temperatures.

Boutique proposals, q

The key determinant of whether an MCMC sampler works well or not is the choice of proposal q.

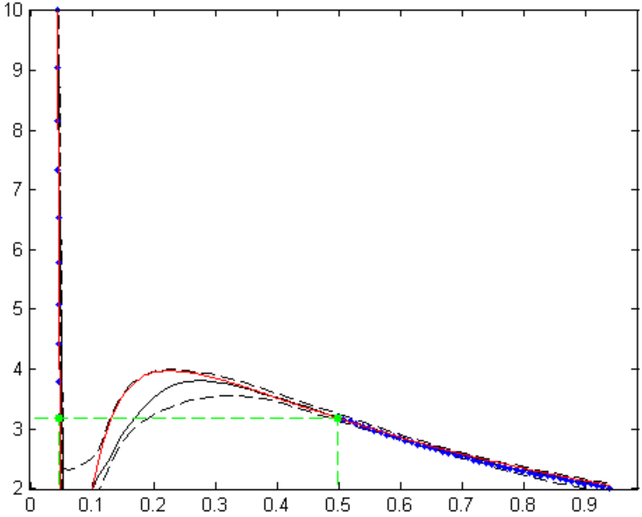
In the GUI we have a carefully tuned proposal that combines multiple different moves.

- With probability p_1 we update a single parameter using a Gaussian random walk proposal
- With probability p_2 we update a block of two parameters using a Gaussian random walk
- With probability p_3 we update all 7 parameters
- With probability p_4 we propose a swap between parameter values in chains of different temperatures.

There are multiple numbers to tune in these proposals:

- The number of different chains and the temperature of each one
- The probabilities of each type of update
- The variance for each type of move, at each temperature

CCS - parallel tempering



Any questions?